

# Measuring Patient Safety in Primary Care: The development and validation of the “Patient Reported Experiences and Outcomes of Safety in Primary Care” (PREOS-PC)

Ignacio Ricci-Cabello (PhD)<sup>1</sup>, Anthony J. Avery (PhD)<sup>2</sup>, David Reeves(PhD)<sup>3</sup>, Umesh T. Kadam (PhD)<sup>4,5</sup>, Jose M. Valderas (PhD)<sup>6</sup>

1. Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK.
2. Division of Primary Care, School of Medicine, University of Nottingham Medical School, Queen's Medical Centre, Nottingham, UK.
3. Centre for Biostatistics, Institute of Population Health, University of Manchester, Manchester, UK.
4. Arthritis Primary Care Research Centre, Institute for Primary Care and Health Sciences, University of Keele, Keele, UK.
5. Health Services Research Unit, Institute for Science and Technology in Medicine, University of Keele, Keele, UK.
6. Health Services & Policy Research Group, Patient Centred Care, University of Exeter Collaboration for Academic Primary Care (APEX), University of Exeter Medical School, University of Exeter, Exeter, UK.

Corresponding author: Jose M. Valderas. University of Exeter Medical School, Smeall Building Luke's Campus, EXETER, EX1 2LU. [J.M.Valderas@exeter.ac.uk](mailto:J.M.Valderas@exeter.ac.uk). Tel: 0044 01392 722743

Support: This research is part-funded by the National Institute for Health Research School for Primary Care Research (NIHR SPCR). The views expressed are those of the authors and not necessarily those of the NIHR, the NHS or the Department of Health.

Word count: 2,951

Figure: 1; Tables: 4; Appendix: 1; Online Appendices: 5

## ABSTRACT

**Purpose:** To develop and validate a patient reported instrument for the measurement of patient safety related experiences and outcomes in Primary Care.

**Method:** The instrument was developed in a multistage process supported by an international expert panel and informed by a systematic review of instruments, a meta-synthesis of qualitative studies, four patient focus groups, 18 cognitive interviews and a pilot study. The trial version of Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC) covered five domains (11 scales): practice activation (1); patient activation (1); experiences of patient safety (1); harm (6); and general perceptions of patient safety (2). The questionnaire was posted to 6,736 patients in 45 practices across England. We used “gold standard” psychometric methods to evaluate its acceptability, reliability, structural and construct validity, and ability to discriminate among practices.

**Results:** 1,244 (18.4%) completed questionnaires were returned. Median item-specific response rate was 91.3% (interquartile range 28.0%). No major ceiling/floor effects were observed. All six multi-item scales showed high internal consistency (Cronbach's  $\alpha$  0.75-0.96). Factor analysis, correlation between scales, and known group analyses generally supported structural and construct validity. The scales demonstrated a heterogeneous ability to discriminate between practices. The final version of PREOS-PC consisted in five domains, eight scales and 58 items.

**Conclusions:** PREOS-PC is a new multi-dimensional patient safety instrument for primary care developed with experts and patients. Initial testing shows its potential for use in primary care and future developments will further address its use in actual clinical practice.

**Key words:** Patient Safety; Primary Care; Patient-Centered Care; Health Care Evaluation Mechanisms; Health Care Surveys

**Abbreviations:** CFI, comparative fit index; ICC, intra-class correlations; PREOS-PC, Patient Reported Experiences and Outcomes of Safety in Primary Care; SRMR, standardized root-mean residual.

## INTRODUCTION

Patient safety, defined by the World Health Organization as “*the prevention of errors and adverse effects to patients associated with health care*”,<sup>1</sup> is a growing interest in primary care systems.<sup>2</sup>

Despite the potential impact on population health, major gaps remain in the understanding of primary care patient safety particularly due to the lack of appropriate measurement methods,<sup>2</sup> limiting our ability to obtaining reliable and repeatable rates of events for safety improvement, and for research to identify fundamental underlying causes and mechanisms.

Current tools rely almost exclusively on information supplied by healthcare providers (e.g. safety culture questionnaires, or voluntary reporting of safety events)<sup>3</sup>. A growing body of evidence, however, suggests that patients are sensitive to and able to recognise a range of problems in healthcare delivery,<sup>4,5</sup> which are not identified by traditional systems of healthcare monitoring.<sup>6,7</sup> Patient reports constitute a reliable source of information<sup>8,9</sup> and have potential to improve the systematic detection of problems in healthcare.<sup>10-13</sup>

Our recent systematic review of primary care patient reported safety measures showed that such instruments largely focus on a small number of relevant dimensions, mostly related to medication problems, and do not allow for a comprehensive assessment of care safety.<sup>14</sup>

We aimed therefore to develop a patient-reported instrument for comprehensively measuring experiences and outcomes of patient safety in primary care, and to test its psychometric properties.

## **METHODS**

Based on quality standards for instrument development and evaluation,<sup>15</sup> the following steps were followed in the development of the new measure: (i) deriving framework for questionnaire domains based on the literature and expert consensus, (ii) identifying and piloting relevant domains and items; and (iii) psychometric testing, including acceptability, internal consistency, construct validity and response bias.

### **Conceptual framework**

Two members of the research team supported by two external experts (see Acknowledgements) reviewed and discussed the conceptual models proposed for patient safety in primary care.<sup>1,16-22</sup> Consensus emerged on three necessary elements for patients safety events: 1) patient interaction with the healthcare system (including self-management); 2) standards of care (with failure to adhering to them possibly due to error or mistake, but also due to other causes); and 3) actual or potential harm to patients, conceptualized as deterioration in health (including physical, mental and social well-being) . An event was hence defined as “Harm or potential harm to one or more patients either due to an interaction with the healthcare system that fails to adhere to accepted standards of care (attributable to error or systemic dysfunctions), or to the intrinsic risks of healthcare (interventions).”

We extracted domains from a meta-synthesis of qualitative studies on patients’ experiences and perceptions of patient safety in general practices: factors contributing to safety events; experiences of safety events (active failures and harm); and patient and provider responses to safety events.<sup>23</sup> Additional domains and themes were obtained through four focus groups with 27 primary care users,<sup>24</sup> and from 23 instruments identified in our previous systematic review.<sup>14</sup>

After removing redundant domains and combining overlapping ones, five main domains emerged: practice activation (what does the practice do to create a safe environment); patient activation (how pro-active is the patient in relation to their safety); experiences of patient safety

events (errors); outcomes of patient safety (harm), and overall perceptions of patient safety (how safe do patients perceive their practice).

### **Identification of items and instrument refinement**

An expert committee composed of five international experts in patient safety in primary care, three local experts and two members of the public (see Acknowledgements) was convened to support the development of the questionnaire (Figure 1).

[Figure 1 about here]

Items were extracted from previous instruments<sup>14</sup> to generate an item pool, further populated with items proposed by the development team based on the literature reviews and the focus groups. Response scales were homogenized wherever feasible. A first draft of the questionnaire was produced and then revised an iterative process (four iterations over 12 months) informed by repeated feedback from the expert committee.

Four waves of cognitive testing (think-aloud technique) were undertaken, including thirteen individual interviews (lasting 45-60 minutes), carried out with members of the public purposefully selected to represent a range of socio-demographic backgrounds.<sup>25</sup>

In a pilot with 1,975 patients in 26 English general practices the feasibility of administration of a pre-trial version of the instrument was tested, and the information was also used in an additional round of expert committee feedback and cognitive interviews (five).

## Psychometric Evaluation

The trial version of the questionnaire was sent in June 2014 to 6,736 patients registered in 45 practices purposefully sampled to ensure maximal variation (list size, levels of deprivation) and distributed across five regions in the North, Centre and South of England. Practices sent the questionnaire to a computer generated random sample of 150 adult (aged  $\geq 18$ ) patients with at least one contact with their practice in the last 12 months. Due to funding constraints a reminder was feasible only for ten practices, and it was sent after an interval of approximately two weeks.

Information on practice characteristics is available in Online Appendix 1. Practices were asked to complete the tool PC SafeQuest,<sup>26</sup> a measure of healthcare professionals' perceptions of the practice's safety climate. Ethical approval was granted by Nottingham Research Ethics Committee (Reference 13/EM/0258; July 2013).

The acceptability of the questionnaire was evaluated through examination of individual item response rates. Scale scores were calculated as the percentage of the maximum score achievable on all items, with scores ranging from 0 (very dissatisfied/totally disagree, etc.) to 100 (very satisfied/ totally agree, etc.). Where responses were missing for 50% or more of the items in a scale, it was scored as missing; otherwise a score was derived using the available items without any imputation.

Internal consistency was examined using inter-item correlations (coefficients  $\geq 0.3$ )<sup>27</sup> and Cronbach's- $\alpha$  ( $\alpha \geq 0.7$ ).<sup>28</sup> Test-retest reliability was analysed using one way random effects intra-class correlations (ICC), (ICC  $\geq 0.7$ ) with data from a sample of 235 respondents who had been invited to complete it twice approximately two weeks apart.<sup>15</sup>

Confirmatory factor analysis was conducted to examine the construct validity of the pre-hypothesized scales. Goodness-of-fit statistics examined included the Satorra-Bentler Chi-squared statistic, comparative fit index (CFI), and standardised root-mean residual (SRMR). We used Hu and Bentler's recommendation for model evaluation<sup>29</sup> consisting in the use of a combinational rule CFI  $> 0.95$  and SRMR  $< 0.09$ . Construct validity was further examined by means of: 1) pre-specified group differences, testing if mean scores discriminated amongst defined groups of: a) users in line

with hypothesised differences (age, ethnicity, language, country of origin, number of long-term conditions and of medications) and of b) practices (list size, deprivation, proportion of patients aged  $\geq 65$ , and safety climate (PC-SafeQuest)), and ; 2) observed correlations amongst PREOS-PC scales with a priori hypothesised relationships.

In order to examine the performance of each scale as a measure of safety at the practice level, we calculated the standard error of a practice mean score as a measure of precision of measurement, and the reliability coefficient (based on the between-practice intra-cluster correlation coefficient) as a measure of ability to discriminate between practices. Both measures are influenced by sample size: we based them on the mean number of patient per practice, but also estimated the sample size required to achieve reliable discrimination between practice scores at the 0.7 level.

Finally, *post-hoc* sensitivity analyses were carried out to examine the magnitude of potential response bias. In the subgroup of practices where reminders were sent, we used hierarchical regression models (adjusting for clustering effect) to compare patient characteristics and scale scores between patients responding to initial invitations vs those responding to reminders. In order to account for skewed score distributions bootstrap methods (50 samples) were used.

All data manipulation and analysis was conducted using STATA version 12.0.

## RESULTS

### PREOS-PC

The resulting instrument ‘Patient Reported Experiences and Outcomes of Safety in Primary Care’ (PREOS-PC) invites patients to report on their perceptions and experiences in relation to the safety of the healthcare received in their GP practice over the past 12 months (Table 1). The trial version contained 54 standardized items and seven open ended questions. Forty-two standardized items were distributed across eleven scales covering all five domains. The remaining twelve standardized items captured details on a specific event (where did the event occur; what actions were taken, etc.) and therefore were not part of any scale since their purpose was descriptive rather than evaluative.

[Table 1 about here]

### Response rate

The overall response rate was 18.4% (1,244/6,736), an average of 28 per practice. The response rate for patients who received a reminder (29.6% (354/1,195)) almost doubled that of patients who did not receive it (16.1% (890/5541)).

Compared to the overall characteristics of all eligible patients registered in the 45 participating practices, respondents were more likely to be female (59% versus 51%), aged  $\geq 65$  (39% versus 19%) and of ‘‘white’’ ethnicity (91% versus 82%) (Table 2). In our sensitivity analyses comparing demographic characteristics and scale scores between patients responding to initial invitations vs those responding to reminders, we observed that the youngest and oldest age-groups and those taking less than four medications were less likely to respond to the first mailing (Online Appendix 2). However no differences in scores between those two groups were observed for any of the scales.



[Table 2 about here]

## **Acceptability**

Median item response rate was 91.3% (interquartile range 69.6%-92.4%). When items were ranked according to non-response, all items in the lower quartile pertained to the “experiences of the most recent safety problem” construct.

There was no evidence of significant ceiling or floor effects except for two items: “harm causing increased personal needs” and “harm causing increased financial needs” (80.1% and 80.4% of patients reporting “not at all”, respectively).

## **Reliability**

The six pre-hypothesized multi-item scales demonstrated high internal consistency (Cronbach’s  $\alpha$ : 0.75-0.96) and adequate homogeneity (inter-item correlations: 0.22-0.83) (Table 3). However, test-retest intra-class correlation coefficients were above the 0.7 standard for only two of the eleven scales (practice activation and health domain specific harm).

[Table 3 about here]

## **Practice-level precision and discrimination**

Taking a standard error of 5 percentage points (on the scale of 0 to 100) as indicating good precision, practice mean scores for all the globally-applicable scales, bar Patient Activation, demonstrated high precision. However, practice means on the sub-set of specific scales (i.e. patients who reported harm) showed very low precision (in all cases a standard error of >13 points).

Between-practice ICCs were mostly low ( $<0.03$ ) suggesting that patient scores only weakly clustered within practices. This is reflected in the low reliability coefficients (all  $<0.7$ ), indicating that although precise, the practice mean scores do not discriminate well between practices in terms of patient perceptions of safety. However, for most scales a sample of around 100 patients would be sufficient to produce scores that discriminate well (i.e. with reliability  $\geq 0.7$ ).

## **Validity**

### *Structural validity*

Confirmatory factor analysis was performed on the five multi-item scales with more than two items and provided evidence for high structural validity (Online Appendix 3). Three of the models met Hu and Bentler's criteria, suggesting adequate goodness-of-fit. Moderately high item-total correlations, high internal consistency coefficients, and the results of the factor analysis indicated that each scale measures a single construct, and that the items can be combined to produce summary scores.

### *Construct Validity*

The great majority of pairwise correlations supported our pre-specified hypothesis (Online Appendix 4). Whereas the results from the analyses of hypothesised differences between groups of patients generally supported the construct validity of the scales examined, the results from the analyses based on practice characteristics were largely inconclusive (Table 4).

[Table 4 about here]

## **Further modifications and final version of PREOS-PC**

Final modifications were made to PREOS-PC based on the results of the psychometric analyses (Online Appendix 5). The modifications mostly concerned the three single item scales in the harm domain (“time to recover from overall harm”; “amount of overall harm experienced”; and “impact of overall harm on overall health”). They were removed because they measured very similar constructs to the three multi-item harm-related scales that remained in the questionnaire which demonstrated better psychometric properties. The final version of PREOS-PC includes 58 items and eight scales (Appendix A).

## DISCUSSION

The PREOS-PC instrument has been developed for providing a comprehensive measure of patient-centred evaluations of patient safety in Primary Care, filling a gap identified in a previous systematic review.<sup>14</sup> It has been developed following the highest standards of instrument development and this study provides preliminary evidence supporting its reliability and validity.

### Strengths and limitations

This study presents a number of methodological strengths. Evidence of the content and face validity of PREOS-PC is supported by the development of the conceptual model, the preparatory qualitative work undertaken,<sup>23</sup> a systematic review of instruments,<sup>14</sup> and by the iterative process of questionnaire development, which was supported by an expert committee. The questionnaire content covers all of the key dimensions within our conceptual framework for primary care patient safety. It was piloted in a large sample of adults registered at a wide range of practices across England. Well established procedures for the assessment of patient reported instruments<sup>15</sup> were applied to examine its reliability and validity.

In terms of limitations, our study had a low response rate (18.4%), substantially lower than response rates from similar large scale surveys such the GP Patient Survey<sup>30</sup> (39%). The sub-sample of patients who received a reminder demonstrated a substantially higher response rate (29.6%); it seems reasonable to anticipate that the inclusion of a second reminder (as is the case for the GP Patient Survey) could have increased the response rate even further.

Non-response can constitute a bias since non-respondents might differ from respondents on the key measures of interest. Meta-analyses suggest that provided rigorous probability sampling processes (such as those used in our study) are followed, the association between response rates and non-response bias within samples is generally weak.<sup>31</sup> Our *post-hoc* analyses showed that although low-response rate resulted in an over representation of elderly and polymedicated patients, this did not

affect to the scale scores, suggesting that response bias did not significantly limit our estimations of the psychometric properties of the instrument.

We observed skewed score distributions for a number of items and scales. However skew is common in questionnaires assessing patients' views of medical care<sup>32,33</sup> and does not necessarily limit the ability to reliably distinguish practices and patient subgroups with sufficient sample sizes as ours.<sup>34</sup>

The acceptability of the "Most recent safety problem" section was relatively low, with only 60% of eligible participants adequately completing that section. This could be partially explained by potentially unclear instructions in the branching question preceding that section. This has subsequently been amended to increase clarity. However, it may also suggest that some patients are reluctant to provide what might be considered overly detailed information about the safety problems experienced.

A substantial proportion of the scales included a low number of items, and five of them were based on single items. This constitutes a limitation since short scales usually present lower levels of accuracy and reliability than scales based on higher number of items. Also, test-retest reliability could not be examined for four of the harm scales due to lack of sufficient cases of harm. This has minor implications for the instrument as three of these have been excluded from the final version. Five of the remaining scales demonstrated low levels of test-retest reliability, suggesting that they are not adequately stable over time. This might suggest interpretation issues; further cognitive testing is needed to inform potential item modification.

We computed scale scores for patients responding to more than 50% of scale items. Measurement errors will be somewhat larger for patients close to the 50% threshold; however, a stricter threshold would result in more patients being fully excluded from the calculation of practice-level scores, potentially increasing the error and bias on those scores, particularly if item non-response is related to patient characteristics or experience. We considered 50% to offer a reasonable balance between these two levels of error and bias. Also, analyses of the psychometric properties were not

stratified levels of service use, and therefore we cannot ascertain the extent to which reliability of the scales was influenced by the number of interactions that patients had with their primary care providers.

Finally, some features of the scales are worth noting, namely the extremely high Cronbach  $\alpha$  in “health domain specific harm” (0.96, which may suggest item redundancy); the low inter-item correlation in the “experiences of safety problems” (0.22, which suggests that problems were quite independents among them); and the low test-retest coefficient for “harm -health and personal care, and financial needs” (-0.02, presumably attributable to the low number of patients reporting harm in our retest sample).

### **Future steps**

Further work is needed prior general application of the instrument. Additional developments will include the assessment of the PREOS-PC responsiveness to change (important if the instrument is to be used as an outcome measure in intervention studies). The development of formal methods for the interpretation of the scores is pending, although provider benchmarking may in itself substantially contribute to this aim. In addition, further work comparing levels of patient safety as measured with PREOS-PC against other measures of the concept is still needed to support the validity of the instrument. Although versions of the current length may be appropriate for research purposes, for service improvement shorter versions may present some advantages. Rasch modelling is especially suitable to identify redundant items.<sup>35</sup> This work is currently underway; so it is the examination of the acceptability and validity of alternative methods for administration (online and in the practice). Future steps will also include the translation of PREOS-PC to a number of different languages, and its cross-cultural adaptation and validation.

### **Conclusions**

PREOS-PC provides a comprehensive measure of patient reported experiences and outcomes of safety in Primary Care. Results from psychometric analysis support its internal consistency and validity, though findings for test-retest reliability were mixed. Further work is needed prior general application of the instrument.

## ACKNOWLEDGMENTS

This research is part-funded by the National Institute for Health Research School for Primary Care Research (NIHR SPCR). The views expressed are those of the authors and not necessarily those of the NIHR, the NHS or the Department of Health.

We would like to thank the following participants and organizations involved in the development of the questionnaire:

Dr Itziar Larizgoitia (WHO), Prof Tony Avery (University of Nottingham), Prof Stephen Campbell (University of Manchester), Prof Charles Vincent (University of Oxford), Dr Angela Coulter (University of Oxford), Dr Sarah P Slight (University of Durham), Dr Umesh Kadam (University of Keele and previously Arthritis Research UK Primary Care Centre), Ms Liz Taylor (Association against Medical Accidents (AvMA)), Mr Derek Shaw, and Mr Antony Chuter (members of the public) for their participation as members of the Expert Committee.

Dr Daniela Goncalves Bradley and Dr Suzanne Shale (Oxford University) for their contribution in the development of the conceptual framework of patient safety.

Mr Antony Chuter (member of the public) and Ms Liz Taylor (Association against Medical Accidents (AvMA)) for their support in recruiting members of the public for the focus groups.

Ms Kate Marsden (University of Nottingham), Ms Katherine Perryman (University of Manchester), Ms Jane Barnett (University of Southampton), Mr Ian Litchfield (University of Birmingham), Ms Sally Thomas (University of Keele) and all the health professionals in the 45 practices involved in this project for distributing the questionnaires to the patients as part of its pilot-testing.

Dr Brian Bell for providing the data on the characteristics of the practices.

Finally we would like to thank all the patients and members of the public that participated in the cognitive interviews, focus groups and completing the survey.



## **Conflict of interest**

The authors declare no conflict of interest.

## REFERENCES

1. World Health Organization. Patient safety. <http://www.euro.who.int/en/health-topics/Health-systems/patient-safety/patient-safety>, 19 August 2015.
2. Lorincz CY, Drazen E, Sokol PE, Neerukonda KV, Metzger J, Toepp MC, et al. Research in Ambulatory Patient Safety 2000–2010: A 10-Year Review. Chicago IL: 2011. [https://c.ymcdn.com/sites/npsf.site-ym.com/resource/resmgr/PDF/Research-in-Amb-Pat-Saf\\_AMAr.pdf](https://c.ymcdn.com/sites/npsf.site-ym.com/resource/resmgr/PDF/Research-in-Amb-Pat-Saf_AMAr.pdf), 19 August 2015.
3. Spencer R, Campbell S. Tools for primary care patient safety: a narrative review. *BMC Fam Pract.* 2014;15:166.
4. Weingart SN, Pagovich O, Sands DZ, Li JM, Aronson MD, Davis RB, et al. Patient-reported service quality on a medicine unit. *Int J Qual Health Care.* 2006;18:95-101.
5. Schwappach DL. Review: engaging patients as vigilant partners in safety: a systematic review. *Med Care Res Rev.* 2010;67:119-48.
6. Levtzion-Korach O, Frankel A, Alcalai H, Keohane C, Orav J, Graydon-Baker E, et al. Integrating incident data from five reporting systems to assess patient safety: making sense of the elephant. *Jt Comm J Qual Patient Saf.* 2010;36:402-10.
7. Evans SM, Berry JG, Smith BJ, Esterman AJ. Consumer perceptions of safety in hospitals. *BMC Public Health.* 2006;6:41.
8. King A, Daniels J, Lim J, Cochrane DD, Taylor A, Ansermino JM. Time to listen: a review of methods to solicit patient reports of adverse events. *Qual Saf Health Care.* 2010;19:148-57.
9. Bjertnaes O, Deilkas ET, Skudal KE, Iversen HH, Bjerkan AM. The association between patient-reported incidents in hospitals and estimated rates of patient harm. *Int J Qual Health Care.* 2015;27:26-30.
10. Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. *BMJ Qual Saf.* 2014;23:678-89.

11. Weingart SN, Pagovich O, Sands DZ, Li JM, Aronson MD, Davis RB, et al. What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med.* 2005;20:830-6.
12. Lawton R, O'Hara JK, Sheard L, Reynolds C, Cocks K, Armitage G, et al. Can staff and patient perspectives on hospital safety predict harm-free care? An analysis of staff and patient survey data and routinely collected outcomes. *BMJ Qual Saf.* 2015;24:369-76.
13. Donaldson LJ. The wisdom of patients and families: ignore it at our peril. *BMJ Qual Saf.* 2015. [Epub ahead of print]
14. Ricci-Cabello I, Goncalves DC, Rojas-Garcia A, Valderas JM. Measuring experiences and outcomes of patient safety in primary care: a systematic review of available instruments. *Fam Pract.* 2015;32:106-19
15. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health.* 2008;11:700-8.
16. Council of the European Union. European Council recommendation on patient safety, including the prevention and control of healthcare associated infections. Brussels, 2009. [http://ec.europa.eu/health/ph\\_systems/docs/patient\\_rec2009\\_en.pdf](http://ec.europa.eu/health/ph_systems/docs/patient_rec2009_en.pdf), 19 August 2015.
17. Kohn LT, Corrigan JM, Donalson MS. To err is human: building a safer health system. Washington DC; 2000.
18. Agency for Healthcare Research and Quality. AHRQ PSNet patient safety network. PSNet Glossary Rockville, MD: Agency for Healthcare Research and Quality; 2011. <http://psnet.ahrq.gov/glossary.aspx>., 19 August 2015.
19. Henriksen K, Battles JB, Keyes MA, Grady ML. *Advances in patient safety: new directions and alternative approaches.* Agency for Healthcare Research and Quality Rockville; 2008.
20. Vincent CA. *Patient Safety.* 2 ed. Oxford: Wiley Blackwell; 2010.

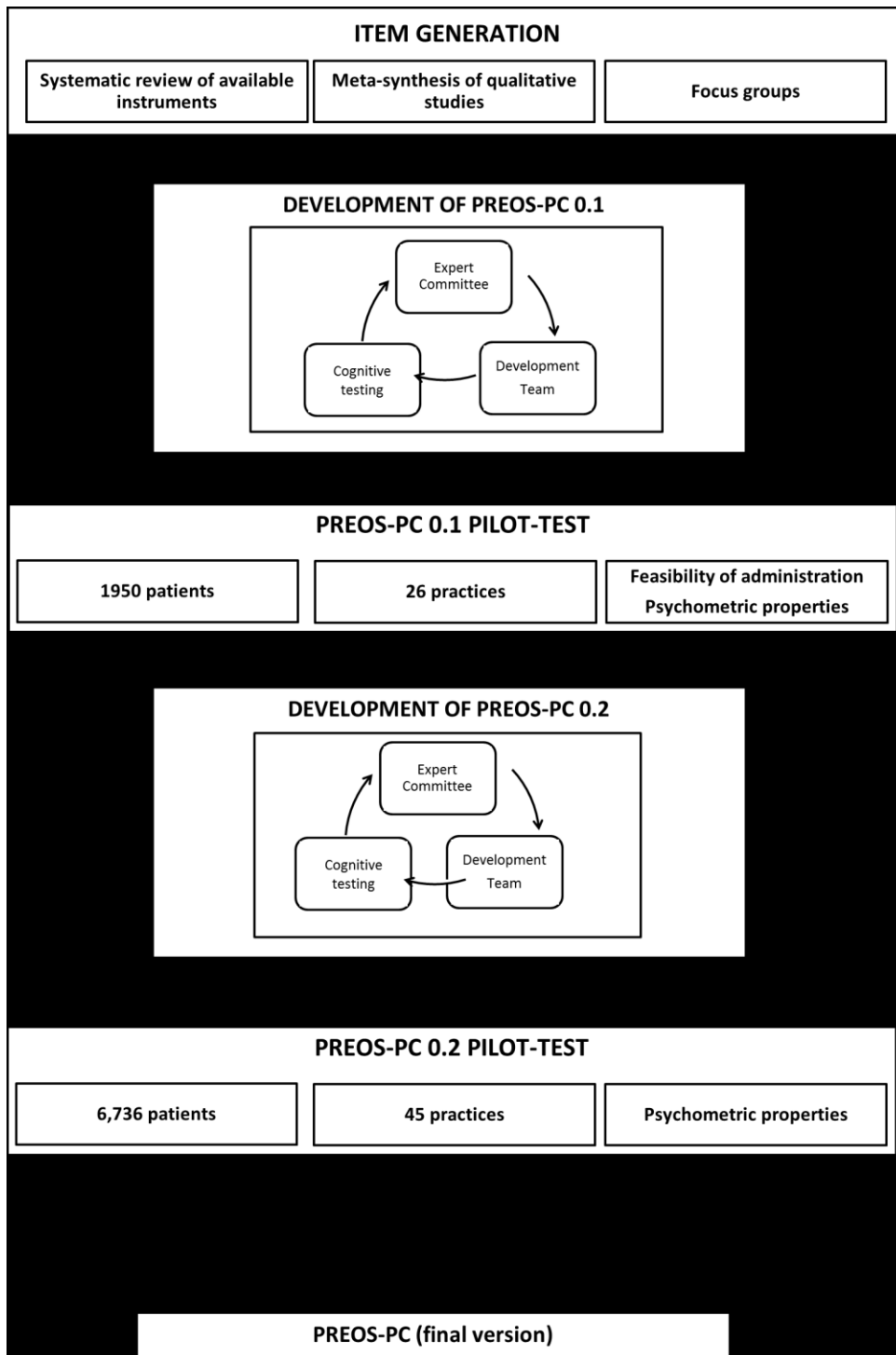
21. Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an International Classification for Patient Safety: key concepts and terms. *Int J Qual Health Care*. 2009;21:18-26.
22. Improving Chronic Illness Care. The Chronic Care Model. [http://www.improvingchroniccare.org/index.php?p=Model\\_Elements&s=18](http://www.improvingchroniccare.org/index.php?p=Model_Elements&s=18), 19 August 2015.
23. Ricci-Cabello I, Goncalves DC, Campbell S, Slight S, Valderas JM. Patients' Experiences of Patient Safety in Primary Care in England: A Systematic Review and Meta-Synthesis. 42 North American Primary Care Research Group (NAPCRG) Annual Meeting; New York 2014.
24. Ricci-Cabello I, Berenguera A, Pujol-Ribera E, Pons-Vigués M, Valderas JM. [Exploring patients' perceptions of patient safety in primary care in England: a qualitative study]. *Gac Sanit*. 2014;28:122-292.
25. Ericsson K A. *Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks*. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press; 2006.
26. De Wet C, Spence W, Mash R, Johnson P, Bowie P. The development and psychometric evaluation of a safety climate measure for primary care. *Qual Saf Health Care*. 2010;19:578-84.
27. Streiner DL, Norman GR. *Health Measurement Scales: A practical guide to their development and use*. Oxford: Oxford University Press; 2008.
28. Nunnally JC, Bernstein IH. *Psychometric theory* (3rd Edition). New York: McGraw-Hill; 1994.
29. Hu L-t, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*. 1998;3:424-53.
30. Campbell J, Smith P, Nissen S, Bower P, Elliott M, Roland M. The GP Patient Survey for use in primary care in the National Health Service in the UK--development and psychometric characteristics. *BMC Fam Pract*. 2009;10:57.
31. Groves RM. Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*. 2006;70:646-75.

32. Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res.* 2009;44:501-18.
33. Mead N, Bower P, Roland M. The General Practice Assessment Questionnaire (GPAQ) – Development and psychometric characteristics. *BMC Fam Pract.* 2008;9:3.
34. Elliott MN, Haviland AM, Kanouse DE, Hambarsoomian K, Hays RD. Adjusting for subgroup differences in extreme response tendency in ratings of health care: impact on disparity estimates. *Health Serv Res.* 2009;44:542-61.
35. Tennant A, McKenna SP, Hagell P. Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value Health.* 2004;7:S22-6.

**Appendix A.** Final version of the PREOS-PC questionnaire

**Figure 1:** Development process of the Patient Reported Experiences and Outcomes of Patient Safety in Primary Care (PREOS-PC)

**Figure 1:** Development process of the Patient Reported Experiences and Outcomes of Patient Safety in Primary Care (PREOS-PC)



**Table 1.** Structure of the trial version of the Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC)

Domain	Quantitative summary		Open ended-questions
	Constructs	Items in scale (n) *	
Practice activation	Practice activation	11	1
Patient activation	Patient Activation	2	1
Experiences of safety problems	Types of patient safety problems experienced	12	3
	Most recent experience: type of patient safety problem; location; people involved and degrees of responsibility (including patient); preventability	N/A (12 items)	
Outcomes of patient safety (harm)	Health domain specific harm	5	2
	Health and personal care, and financial needs	3	
	Time to recover from harm (type specific)	4	
	Time to recover from harm (overall)	1	
	Amount of harm experienced (overall)	1 †	
	Impact on overall health	1	
General perceptions of patient safety	Trustworthiness	1	0
	Overall rating of patient safety	1 †	
Total	12 constructs	11 scales (54 items)	7

\* Items are based on Likert scale, unless otherwise stated. † Visual Analogue Scale.

N/A: not applicable (no evaluation scale hypothesised)



**Table 2.** Demographic characteristics of the participants

	N (%)
Sex <sup>1</sup>	
Male	497 (41.11%)
Female	712 (58.89%)
Age <sup>2</sup>	
18-34	140 (12.03%)
35-64	570 (48.97%)
≥65	454 (39.00%)
Ethnicity <sup>3</sup>	
White	1082 (91.15%)
Other ethnic group	105 (8.85%)
Educational level	
Degree, degree equivalent and above	411 (35.16%)
Other qualifications	532 (45.51%)
No qualifications	226 (19.33%)
Health status	
Very good/ Good	892 (73.54%)
Fair /Bad /Very bad	321 (26.46%)
Number of long term conditions	
0	330 (27.99%)
1	329 (27.91%)
2-3	366 (31.04%)
>3	154 (13.06%)
Number of medications taken	
0	344 (30.10%)
1-2	311 (27.21%)
3-4	222 (19.42%)
>4	266 (23.27%)

<sup>1</sup>Mean (SD) proportion of female registered in the 45 practices that participated in the study: 0.51 (0.05).

<sup>2</sup>Mean (SD) proportion of eligible patients aged >65 registered in the 45 practices that participated in the study: 0.20 (0.01).

<sup>3</sup>Mean (SD) proportion of patients from non-white ethnicity registered in the 45 practices that participated in the study: 0.18 (0.04)

**Table 3.** Distribution of scores and reliability of the Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC) scales

Domain	Construct	N	Observed range		Mean (SD)	Respondents with lowest possible score (%)	Respondents with highest possible score (%)	Internal consistency		Test-retest reliability *	Practice mean scores			
			Min	Max				Cronbach's $\alpha$ †	Inter-item correlation †		ICC** (95%CI)	Precision (standard error)	Intra-cluster correlation coefficient (95%CI)	Reliability
Practice activation	Practice activation	1132	9.09	100	83.69 (18.01)	0	20.56	0.89 (0.86; 0.90)	0.41 (0.39; 0.43)	0.72 (0.55; 0.83)¶	3.44	0.07 (0.03; 0.11)¶	0.66	31
Patient Activation		966	0	100	25.1 (30.8)	47.0	5.58	0.80	0.67	0.55 (0.25;0.75)¶	6.57	0.022 (0.001; 0.050)¶	0.33	104
Experiences of safety problems		1171	0	72.72	4.8 (9.4)	63.19	0	0.75 (0.71; 0.76)	0.22 (0.20; 0.24)	0.57 (0.37; 0.72)	1.83	0.02 (0.00; 0.05)	0.36	109
Harm	Health domain specific harm	1053	0	100	4.7 (14.1)	81.8	0.29	0.96 (0.95; 0.96)	0.83 (0.82; 0.85)	0.72 (0.55;0.83)¶	2.88	0.025 (0.001; 0.053)¶	0.38	90
	Health and personal care, and financial needs	1043	0	91.66	2.4 (10.8)	92.6	0	0.88 (0.78; 0.89)	0.72 (0.63; 0.80)	-0.02 (-0.29; 0.26)	2.23	0.019 (0.001; 0.046)¶	0.31	118
	Time to recover from	157	0	75	24.7 (26.7)	37.7	0	0.81 (0.71; 0.86)	0.52 (0.45; 0.67)	Not estimated‡	13.59	0.057 (0.001; 0.21)	0.18	309 <sup>§</sup>

	harm (type specific)													
	Time to recover from harm (overall)	162	0	100	56.4 (41.6)	27.6	32.27	N/A	N/A	Not estimated‡	19.50	0.17 (0.001; 0.34)¶	0.45	85 <sup>§</sup>
	Amount of (overall) harm experienced	169	0	100	35.3 (26.3)	3.0	0.60	N/A	N/A	Not estimated‡	13.57	0.000 (0.000; 0.126)	-	-
	Impact on overall health	168	0	100	25.4 (32.5)	54.2	7.74	N/A	N/A	Not estimated‡	15.55	0.11 (0.001; 0.256)¶	0.32	148 <sup>§</sup>
General perceptions of patient safety	Trustworthiness	1133	0	100	87.5 (16.1)	0.3	56.78	N/A	N/A	0.26 (-0.02;0.50)¶	3.16	0.032 (0.002; 0.061)¶	0.45	71
	Overall rating of patient safety	1139	0	100	86.0 (16.8)	0.2	19.79	N/A	N/A	0.24 (-0.03; 0.48)¶	3.29	0.029 (0.001; 0.058)¶	0.43	78

† Mean (max, min); \* Based on data from 64 patients that completed the questionnaire again after two week (not estimated when eligible respondents were <10). \*\* One way random effects intra-class correlation coefficients and 95% confidence intervals. ‡Not enough data to conduct the analyses (less than 10 of the respondents completing the retest questionnaires reported harm experiences); ¶  $p < 0.05$ ; §Responses needed to have sufficient cases reporting harm; N/A, non-applicable (single item scales).

**Table 4.** Known Group Analysis based on patients’ and practices’ characteristics

	Practice activation*	Experiences of Safety Problems**	Impact on health (health domain specific)**	Impact on health and personal care, and financial needs **	Overall rating of patient safety *
<b>A. Patients’ characteristics</b>	<b>β (95%CI)<sup>a</sup></b>	<b>β (95%CI)<sup>a</sup></b>	<b>β (95%CI)<sup>a</sup></b>	<b>β (95%CI)<sup>a</sup></b>	<b>β (95%CI)<sup>a</sup></b>
Number of long term conditions	-0.13 (-0.87; 0.60)	0.49 (0.53; 0.93)	1.58 (0.82;2.35)	0.88 (0.36;1.40)	0.11 (-0.42;0.64)
Number of medications	0.03 (-0.34;0.39)	0.14 (-0.6;0.33)	0.65 (0.13;1.16)	0.49 (0.19; 0.80)	0.17 (-0.17;0.53)
English as a second language					
Yes (n=87)	1	1	1	1	1
No (n=1,118)	5.71 (0.57;10.86)	-4.37 (-8.30;-0.44)	-2.83 (-7.65; 1.98)	7.87 (2.95; 12.80)	-4.67 (-10.36;1.02)
Born in the UK					
Yes (n=1,093)	1	1	1	1	1
No (n=112)	-2.46 (-6.95; 2.03)	2.83 (-0.02;5.68)	1.48 (-2.38;5.34)	-5.13 (-8.64; -1.63)	3.13 (-0.19;6.47)
<b>B. Practices’ characteristics</b>	<b>r (95%CI)<sup>b</sup></b>	<b>r (95%CI)<sup>b</sup></b>	<b>r (95%CI)<sup>b</sup></b>	<b>r (95%CI)<sup>b</sup></b>	<b>r (95%CI)<sup>b</sup></b>
Practice list size	0.24 (-0.06; 0.50)	-0.27 (-0.52,0.03)	-0.11 (-0.40; 0.19)	-0.02 (-0.31; 0.28)	0.01 (-0.20;0.39)
Practice proportion of patients aged> 65	0.06 (-0.25;0.35)	-0.08 (-0.37,0.22)	-0.07 (-0.36; 0.24)	-0.04 (-0.34; 0.26)	0.14 (-0.16;0.42)
Deprivation score <sup>c</sup>	-0.32 (-0.56;-0.02)	0.30 (0.00;0.55)	0.20 (-0.10; 0.47)	0.18 (-0.13; 0.45)	-0.36 (-0.60;-0.07)
Safety Climate <sup>d</sup>	-0.09 (-0.45;0.29)	0.06 (-0.32,0.42)	0.10 (-0.28; 0.46)	0.15 (-0.50; 0.25)	0.07 (-0.31;0.43)

a: bivariate linear regression; b: correlation coefficient; c: deprivation measured using the Index of Multiple Deprivation, which higher scores indicating higher deprivation levels; d, Safety climate measured using the instrument “PC-SafeQuest” (information only available from 31 practices); \* Higher scores indicate perception of safer practices. \*\* Higher scores indicate more severe/frequent access problems or more severe harm.